METHOD FOR CHARACTERIZING SAMPLES

The present invention relates to methods for characterizing samples. These methods are i.a. used to investigate test samples from a production, patients or samples collected in any
5   other way.

Background of the invention

When a sample is to be characterized for components, the components are generally separated from each other in a first step in order to identified and quantified in a later stage.
10   However, it is not always possible to separate the components or it may not be motivated from a time/cost benefit reason. The samples may then be characterized spectroscopically whereby the components are identified by means of their unique spectral responses.

If one has a collection of samples and is aware of which components they comprise, it is, as
15   a rule, trivial to determine their concentrations spectroscopically. This is due even if the spectral responses of the components overlaps each other. If, however, the components are unknown, the problem is muck more complicated . The situation was analysed for the first time in detail by the mathematics Lawton and Sylvestre (Technometrics, 13, 617, (1971)), who showed that it is impossible to find an unique solution even for a 2-component system.
20   In 1990 we developed an experimental method, which partly solved this problem (Kubista, Chemometrics and Intelligent Laboratory Systems, 7, 273, (1990)). We then showed that if one carried out two spectroscopic measurements on each sample, in stead of one as previously used, and the measurements were such that the contribution of the components to these measurements had the same distribution of the intensities, but of different magnitude,
25   then both the spectral responses as well as the concentrations of the components could be determined. Mathematically, these measurements are described using the equations:

$$A = CV \text{ or } a_j(\lambda) = \sum_{i=1}^{r} c_{ij} v_i(\lambda) \qquad j = 1, 2 \dots n$$

30   $$B = CDV \text{ or } b_j(\lambda) = \sum_{i=1}^{r} c_{ij} d_j v_i(\lambda) \qquad j = 1, 2 \dots n$$

wherein A is a matrix comprising spectra of the first type measured on the n samples; B is a matrix comprising spectra of the second type measured on the same n samples; C is a

2

matrix comprising the concentrations of the r different components in the n samples; **V** is a
matrix comprising the normalized spectra of the components; and **D** is a diagonal matrix,
the **r** diagonal elements of which being the ratios between the responses of the components
obtained in the two measurements. All spectra are digitalized in m points. We showed that

5   the concentrations of the components (**C**), their normalized spectral responses (**V**) and the
ratio between their responses obtained in the two measurements (**D**) could be determined
only outgoing from the information obtained from the spectra as measured (**A** and **B**). We
further described how the number of components of the samples (r) could be estimated.

10 One restriction using this method is that the number of components are not allowed to ex-
ceed the number of samples, which from a practical point of view means that the method
can not be utilized on smaller series of samples and can not be applied on the whole for
analysing isolated samples.

15 Several spectroscopic techniques, such as fluorescence, nmr, etc., can generate 2-dimen-
sional data described by the equation:

$$I(\alpha,\beta) = \kappa \sum_{i=1}^{r} I_i(\alpha) c_i I_i(\beta)$$

where the signal, $I(\alpha,\beta)$, is determined as a function of two variables, α and β, and are the

20 sum of the contribution of the components in each point, which contribution is proportional
to their concentrations (c_i) and the products of their (normalized) 1-dimensional responses,
$I_i(\alpha)$ and $I_i(\beta)$. Out of these responses the components can be identified. In a steady state
fluorescence spectroscopy $I_i(\alpha)$ and $I_i(\beta)$ are the excitation- and emissions spectra of the
components and are, as a rule, designated $I_i^{ex}(\lambda_{ex})$ and $I_i^{em}(\lambda_{em})$, wherein λ_{ex} and λ_{em} are the

25 excitation and emission wavelengths. The shape of an excitation spectra of a pure com-
pound is, in general independent of the emission wavelength used at the measurement, and
the corresponding is due for its emission spectrum. The fluorescence signal monitored, if
necessary after a correction for the inner filter effect (Kubista et al, The Analyst, 119, 417
(1994)), is proportional to the concentration of the compound. In a sample containing more

30 compounds the total signal is the sum of the contribution by each component. As fluore-
scence is measured in an arbitrary unit, eq. 1 contains a proportionality constant (κ).

The information of the 2-dimensional spectrum $I(\alpha,\beta)$ is insufficient to unambiguously

3

determine the spectral responses of the components. Different approximative ways have been suggested but these do not function sufficient satisfactorily even for a 2-component mixture (Burdick and Tu, J. Chemometrics, 3, 431, (1989)).

5   The present invention is a method for analysing isolated test samples, or a couple of test samples without using references in such a way that the components can be identified.

Description of the figures.

Figure 1. Emission spectra monitored using different excitation wavelengths using a parallel
10   polarized light (above, left) and a perpendicularly polarized light (above right), respectively. Down to the left the calculated emission spectra of the components are shown, and down to the right the calculated excitation spectra of the components are shown.

Figure 2. A) Excitation spectra registered using different emission wavelengths from two
15   solutions containing POPOP, dimethyl POPOP, antracene, and diphenyl antracene. B) The excitation spectra of the components as calculated.

Figure 3. A) Emission spectra registered using different excitation wavelengths of two solutions containing POPOP, dimethyl POPOP, antracene, and diphenyl antracene. B) The
20   excitation spectra of the components as calculated.

Brief description of the invention

The present invention is a method for analyzing test samples in such a way that its components can be identified without the need for any reference data. The method is based upon
25   the following four steps:

1. The test sample is analyzed using a method generating a 3-dimensional response according to : $I(\alpha,\beta,\gamma) = \sum_{i=1}^{r} \tilde{I}_i\,(\alpha)\tilde{I}_i\,(\beta)\tilde{I}_i\,(\gamma)$,

wherein r is the number of components contributing to the signal, and $\tilde{I}_i(\alpha)$ and $\tilde{I}_i(\beta)$ and
30   $\tilde{I}_i(\gamma)$ are the arbitrarily normalized 1-dimensional responses of the components, which responses normally consist of spectral or concentration variations.

2. The number of components r as the samples contain is estimated.

3. For each component its 1-dimensional responses $I_i(\alpha)$ and $I_i(\beta)$ and $I_i(\gamma)$ are determined.

4

4. Out of the responses, the components are identified.

Detailed description of the present invention

As the title indicates the present invention relates to a method for characterizing isolated test

5   samples in a way that makes it possible to identify its components without any need for

using reference samples. This is done through a strategic design of experiments which

makes it possible to register a 3-dimensional response being proportional to the concentra-

tions of the components, and the contribution from each component is the product of its

specific 1-dimensional responses:

10   $$I(\alpha,\beta,\gamma) = \sum_{i=1}^{r} c_i I_i(\alpha) I_i(\beta) I_i(\gamma)$$

Such registration can be carried using certain forms of fluorescence spectroscopy, e.g., by

means of a time disintegrated monitoring of emission/excitation spectra, i.e., the signal is

registered as a function of excitation wavelength, emission wavelength, and time:

15   $$I(\lambda_{ex},\lambda_{em},t) = \sum_{i=1}^{r} c_i I_i(\lambda_{ex}) I_i(\lambda_{em}) I_i(t)$$

In these cases it is often suitable to gather the concentration of the components ci and the

time declinations to a time dependent concentration:

$$I(\lambda_{ex},\lambda_{em},t) = \sum_{i=1}^{r} c_i(t) I_i(\lambda_{ex}) I_i(\lambda_{em})$$

20   The time can be time after light pulse (whereby $c_i(t)$ is proportional to the fluorescence

declination), time after mixing of e.g., a stop-flow experiment (whereby $c_i(t)$ is the variation

of the concentration of component i with time), time after treatment, such a photo bleaching

(selective destruction of certain components using light), chromatographic or other form of

separation, etc. At the analysis of such data the concentration variation of the components

25   are calculated, as well as their excitation and emission spectra. It is of interest to note that

intermediate components which are neither present at the beginning ($c_i(0) = 0$) or at the end

($c_i(\infty) = 0$) of the experiment can be identified from its calculated spectra.

There is a further possibility in varying the polarization of the light:

30   $$I(\lambda_{ex},\lambda_{em},\alpha) = \sum_{i=1}^{r} c_i I_i(\lambda_{ex}) I_i(\lambda_{em}) I_i(\alpha)$$

or, if the phase-modulated light is utilized, the frequency of the modulation:

$$I(\lambda_{ex},\lambda_{em},\nu) = \sum_{i=1}^{r} c_i I_i(\lambda_{ex}) I_i(\lambda_{em}) I_i(\nu)$$

5

etc.

There is further a possibility in varying the outer parameters which influences the concentrations of the components, such as temperature (pressure, volume, etc.):

5
$$I(\lambda_{ex}, \lambda_{em}, T) = \sum_{i=1}^{r} c_i(T) I_i(\lambda_{ex}) I_i(\lambda_{em})$$

or outer parameters which influence the intensity of the responses of the components, such as external magnetic fields (electrical fields, etc.):

$$I(\lambda_{ex}, \lambda_{em}, M) = \sum_{i=1}^{r} c_i I_i(\lambda_{ex}) I_i(\lambda_{em}) I_i(M) .$$

10

The spectroscopic technique need not be a fluorescence technique. The method can be carried out using most techniques which generates 3-dimensional responses, e.g., nuclear magnetic resonance spectrometry (NMR) mass spectrometry, etc. It can further be carried out using most techniques generating 2-dimensional responses if the responses of the

15   components influence external parameters. Finally, the method can be used using a technique generating 1-dimensional responses, as well, but then it is necessary that two external parameters are varied simultaneously and that their influence on the responses of the components are independent so that their contribution can be factorized.

20   The invention requests that at least two data points are determined in each of the 3 dimensions, i.e.:

$I_i(\alpha)$ wherein $\alpha_1, \alpha_2 \dots \alpha_n$ $l \geq 2$

$I_i(\beta)$ wherein $\beta_1, \beta_2 \dots \beta_m$ $m \geq 2$

$I_i(\gamma)$ wherein $\gamma_1, \gamma_2 \dots \gamma_l$ $n \geq 2$

25

To determine two data points only in all dimensions are, however, of particular meaning as the tolerance of the responses calculated then as a rule is insufficient to be able to identify the components. On the contrary it is quite excellent to have two data points only in one of the dimensions , e.g, $l=2$ (and $m \gg 2$, and $n \gg 2$). This exhibits the advantage that the nume-

30   rical treatment of data is made easier as the responses of the components can be calculated using fast algorithms such as Procrustes rotation and GRAM (Kubista, Chemometrics and Intelligent Laboratory Systems, 7, 273, (1990); Wilson, Sanches & Kowalski, J. Chemometrics, 3, 493, (1989)). In the general case when all l, n, and m are greater than 2, the

6

solution method is much more complicated and thus considerably more time consuming (Liwo, et al, Computers Chem., 21, 89-91, (1997)). Furthermore, it is quite often of interest to carry out the experiment in such a way that one of m and n are considerably greater than the other. The reason hereto is that it as a rule, is sufficient for the identification of the com-

5 ponents, to determine one of their 1-dimensional responses with a high accuracy.

The invention is not limited to determinations that generates 3-dimensional responses but even responses of a higher order can be used. In general it should be satisfying that the response is linear and that the contribution from each component shall be the product of its

10 1-dimensional responses:

$$I(\alpha,\beta,\gamma,\delta....) = \sum_{i=1}^{r} c_i I_i(\alpha) I_i(\beta) I_i(\gamma) I_i(\delta)....$$

Of course, the higher the dimension is the more time consuming the numerical treatment of

15 the determined data will become. However, with regard to the very fast development within the computer area this will hardly be a practical limitation in the future.

The samples to be analysed shall contain substantially the same components, and these shall be present in different, relative concentrations. The samples are analysed in pair using a 2-

20 dimensional method which provides a response which is proportional to the concentrations of the components and the product of the 1-dimensional responses. This can be expressed

as: $$I^A(\alpha,\beta) = \sum_{i=1}^{r} I_i(\alpha) c_i^A I_i(\beta)$$

25 $$I^B(\alpha,\beta) = \sum_{i=1}^{r} I_i(\alpha) c_i^B I_i(\beta)$$

wherein $I^A(\alpha)$ and $I^B(\beta)$ are spectra of the two samples which in the following will be called A and B, determined as a function of the variables α and β, r is the total number of compo-nents contributing to the spectra, $I_i(\alpha)$ and $I_i(\beta)$ are the normalized 1-dimensional responses of the components, and c_i^A and c_i^B are their concentrations, respectively. In a steady-state

30 fluorescence spectroscopy $I_i(\alpha)$ are the normalized excitation spectra of the components, $I_i^{ex}(\lambda_{ex})$, and $I_i(\beta)$ are their normalized emission spectra, $I_i^{em}(\lambda_{em})$:

$$I^A(\lambda_{ex},\lambda_{em}) = \sum_{i=1}^{r} I_i^{ex}(\lambda_{ex}) c_i^A I_i^{em}(\lambda_{em})$$

7

$$I^{B}(\lambda_{ex},\lambda_{em}) = \sum_{i=1}^{r} I_{i}^{ex}(\lambda_{ex})c_{i}^{B}I_{i}^{em}(\lambda_{em})$$

The information in these spectra is treated in two steps. First the number of components, r,

5 is determined, and then the 1-dimensional responses of the components,

$I_{i}^{ex}(\lambda_{ex})$, and $I_{i}^{em}(\lambda_{em})$.

When r has been determined, the spectral responses of the components.

The equations:

$$I^{A}(\lambda_{ex},\lambda_{em}) = \sum_{i=1}^{r} I_{i}^{ex}(\lambda_{ex})c_{i}^{A}I_{i}^{em}(\lambda_{em})$$

10

$$I^{B}(\lambda_{ex},\lambda_{em}) = \sum_{i=1}^{r} I_{i}^{ex}(\lambda_{ex})c_{i}^{B}I_{i}^{em}(\lambda_{em})$$

can be written in matrix form as:

$A=XC^{A}M$

15 $B=XC^{B}M$

wherein A and B are matrixes comprising the spectra determined, X is a matrix comprising the normalized excitation spectra of the components, M is a matrix comprising their normalized emission spectra, and C^{A} and C^{B} are diagonal matrixes comprising the concentrations of the components. By renormalizing one of X or M, the equation system can be rewritten

20 as:

$A=XM$

$B=XDM$

wherein D is a diagonal matrix comprising the ratios between the concentrations of the components $(D=C^{B}/C^{A})$. Using A and B, X, M and D can be calculated using known

25 methods such as Procrustes rotation (Kubista, Chemometrics and Intelligent Laboratory Systems, 7, 273, (1990); and GRAM (Wilson, Sanches & Kowalski, J. Chemometrics, 3, 493, (1989).

As a summary, the present invention relates to a method for experimentally studying two

30 samples spectroscopically so that the information present in the experimental spectra is sufficient to determine the number of components of the samples (r), their spectral responses of the 1st dimension, $I_{i}(\alpha)$, their spectral responses of the 2nd dimension, $I_{i}(\beta)$, and the ratios between their concentrations (c_{i}^{A}/c_{i}^{B}).

8

The most apparent use of the invention is for the analysis of two samples containing
common components. All components need not be common, but the majority of those con-
tributing spectroscopically should be in common (Booksh & Kowalski, J. Chemotrics, 8,
287, (1994)). The number of components is arbitrary and can exceed 2.

5

Another use of the invention is to characterize single samples by first dividing them into
two part samples containing the ingoing components in different proportions. This can be
accomplished in several ways, e.g., by filtering, extracting, chromatographying dialysing,
centrifuging, precipitating, splitting the sample by means of an electrical field, etc. Alter-

10 natively, the original sample can be used as one sample, and an aliquot thereof, which is
created in such a way that the components are present in other proportions, is used as the
second sample. This aliquot can be obtained by selectively eliminating certain components,
e.g., by means of adsorption, precipitation, freezing, distillation, selective decomposing
(e.g., by light, heat, radio lysis), etc. Another possibility is to create two samples from one,

15 is to change the conditions for the determination, e.g., by changing the temperature, pres-
sure, etc. Separation methods, such as different types of chromatography are of interest, as
the components are separated in space, and one, principally arbitrary number of samples can
be obtained which can be analysed in pair. Using spectroscopic techniques which generates
2-dimensional spectra in a fast way, then, furthermore, the detection can be made on-line.

20

Another use of the invention is to determine the concentrations of the components in one
test sample in relation to a standard sample with a high degree of accuracy. The standard
sample and the test sample are analysed as a pair, and the ratio between the concentrations
of the components is obtained as the diagonal element of the D matrix.

25

2-dimensional spectra wherein one of the dimensions is time, are of particular interest,
whereby time is related to time after a disturbance such as a relaxation time. Today, there
are e.g., fluorescence instruments by means of which one can determine complete spectra as
a function of time after lightening (either directly after lightening using a light pulse, or

30 indirectly using phase modulation technique). This gives using α as time, and β as wave
length, the equation system:

$$I^A(t, \lambda) = \sum_{i=1}^{l} I_i(t) c_i^A I_i(\lambda)$$

9

$$I^B(t,\lambda) = \sum_{i=1}^{r} I_i(t)c_i^B I_i(\lambda)$$

from which r, $I_i(t)$, $I_i(\lambda)$ and (c_i^A/c_i^B) can be determined.

5   Example

The invention will be further illustrated in four examples.

Example 1

A sample is characterized using fluorescence spectroscopy, where excitation wave length,

10   emission wave length, and light polarization are varied (Figure 1). This gives raise to a 3-dimensional spectrum according to:

$$I(\lambda_{ex},\lambda_{em},\alpha) = \sum_{i=1}^{r} c_i I_i(\lambda_{ex}) I_i(\lambda_{em}) I_i(\alpha)$$

In the example 650 different emission wave lengths (m), 11 different excitation wave

15   lengths (n) and 2 different polarizations (α = 0o, called parallel polarization, and α = 90o, called perpendicular polarization) (l), are used. From the response determined, $In(\lambda_{ex}, \lambda_{em}, \alpha)$, first the number of components (r) is estimated to 2 (using a statistic test and a visual inspection of the principal components). Then the component specific responses are calculated. For this purpose one uses the fact that only two data points were registered in one

20   of the dimensions (polarization) and rewrote the 3-dimensional response to two 2-dimensional responses.

$$I(\lambda_{ex},\lambda_{em},\alpha = 0°) = \sum_{i=1}^{r} c_i I_i(\lambda_{ex}) I_i(\lambda_{em}) I_i(\alpha = 0°)$$

$$I(\lambda_{ex},\lambda_{em},\alpha = 90°) = \sum_{i=1}^{r} c_i I_i(\lambda_{ex}) I_i(\lambda_{em}) I_i(\alpha = 90°)$$

25   These can be described using the equation system:

$I^0=X\alpha^0M$

$I^{90}=X\alpha^{90}M$

which can be solved using Procrustes rotation (Kubista, Chemometrics and Intelligent Laboratory Systems, 7, 273, (1990)). This gave the normalized excitation intensities of the com-

30   ponents as matrix $X(I_i(\lambda_{ex}))$, (shown down to the right in Figure 1), the normalized emission intensities of the components as matrix M $(I_i(\lambda_{em}))$ (shown down to the left in Figure 1), and the ratios between the responses of components to light of different polarization

10

From the calculated component specific responses, in particular the emission spectra, the

component could be identified as p-bis[2-(5-phenyloxazolyl)]-benzene (POPOP), and antra-

cene. Finally, by comparing standard spectra of POPOP and antracene the concentrations

could be estimated to some micro molars.

Example 2

Two solutions containing the dye compounds POPOP, dimethyl POPOP, antracene and

diphenyl antracene in different proportions were prepared. On these fluorescence excitation

spectra were monitored at several emission wave lengths. The number of components were

10   determined to 4 using a statistic test, and the excitation spectra of the components (Figure

1), emission intensities and the relation between their concentrations in the two samples

were calculated.

Example 3

15   On the same solutions as in Example 1 the fluorescence emission spectra were monitored

using a number of excitation wave lengths . The number of components was determined to 4

using a statistic test, and the emission spectra of the components (Figure 2), excitation in-

tensities and the relation between their concentrations in the two samples were determined.

20   ### Example 4

Characterization of samples containing the dye compound thiazole orange and the polymer

poly(dG) was made. The samples were analysed in pairs using 2-dimensional fluorescence

spectroscopy. They contains thiazole orange and poly(dG) in the relation

[thiazole orange]/[poly(dG)] of 0.05 and 0.025. Neither poly(dG) nor the dye compound is

25   fluorescentic as such but the fluorescence arises when thiazole orange binds to the polymer.

The samples were analysed in two different ways. In one analysis, the fluorescence exci-

tation spectra were monitored at different emission wave lengths. The number of fluorescent

components were identified to two using statistic tests, and their excitation spectra and

emission intensities were calculated.. In the second analysis, the fluorescence emission

30   spectra were monitored using a number of excitation wave lengths. Once again the number

of components was identified to two, and their emission spectra and excitation intensities

were calculated.